韻律情報に基いたあいづち挿入箇所の推定

野口 広彰[†], 小磯 花絵[††], 福田 泰子[*], 伝 康晴[†]
[†]奈良先端科学技術大学院大学, [‡]ATR 知能映像通信研究所, [*]大阪外国語大学

1 はじめに

あいづちは意味理解や統語解析を必要としない素朴な言語現象のひとつであるが、従来の音声対話システムではまったく考慮されてこなかった。しかし、あいづち挿入のタイミングが不適切であったり、あるいはあいづちをまったく打たないことで話者に不安や緊張を与えることが近年明らかになってきており [6]、より人間に近いインターフェースを実現にはあいづちの挿入メカニズムの解明は避けて通れない問題である。

そこで本研究では、あいづちのようなより人間らしい応答を音声対話システム上で実現することを目指して、人間の発話中の適切な箇所であいづちを自動的に挿入するメカニズムを提案する。また、実現にあたっては既存の技術で機械的に抽出が容易な韻律情報のみを用いる。なお、本研究では基礎資料として、筆者らが作成した自由会話音声対話コーパス [7] を使用する。

2 先行研究

あいづちはいつでも打てるというわけではなく、適切なタイミングで打つことによってその機能が果たされる [2]。そこで、人間があいづちを打つタイミングについての研究がなされている。

メイナードはポーズ節の直前に話者自身があいづち挿入の合図を提供していると考察しており [5]、ほとんどの先行研究もポーズ節末に合図があるという仮定を踏襲している。合図としては、上昇下降イントネーション [3, 4]、上昇イントネーション [3]、下降イントネーション [10, 3] などが挙げられている。また、終助詞、間投助詞などの語彙の情報も挙げられている。

音声対話システムへの応用を目的とした研究はまだ少ないが、Ward と岡登らの研究が挙げられる。Ward は「ピッチがピーク値の 30%を下回る区間が 200ms 続いたら 150ms 後にあいづちを打つ」というヒューリスティクスを考案、これをあいづち応答システムとして実装した [11]。岡登らはあいづち挿入の手がかりを、あいづち直前の 200-400ms のピッチの形に求め、この部分のピッチパターンをテンプレートマッチしてあいづち推定を行なった [8]。

上記の関連研究において提唱されているあいづち挿入箇所の予測ルールはいずれもヒューリスティクスであり、そのために扱える韻律特徴が少ない。そこで本研究ではより多くの韻律特徴をもとにして統計的学習を用いて予測ルールを生成する手法を採る。

3 あいづちコンテクストの同定

3.1 目的

本研究では、音声対話コーパス中のあいづち挿入箇所付近の韻律的特徴を統計的に学習し、あいづちコンテクストに特有の特徴を捉えることを目的としている。そのために学習データとして正例と負例が必要である。学習データの採り方に関して以下の問題がある。

1. コーパス中に現われたあいづちのみを正例として使用するか。
2. コーパス中に現われた全てのあいづちを正例とするか。
3. 正例以外を全て負例とするか。

コーパス中に現われたあいづちコンテクストのみを正例とするか あいづちを任意の応答とする考えが先行研究でも有力であり、挿入できる箇所 (正例コンテクスト) であっても個人の癖や対話への参加度によって必ず挿入するとは限らない。よってコーパス中に記録されているあいづちコンテクストの集合は正例コンテクストの部分集合でしかなく、これらのみを正例として学習した場合、正例の再現率が著しく下がる恐れがある。

コーパス中に現われた全てのあいづちコンテクストを正例とするか 本研究では、コーパス中に現われたあいづちコンテクストのうち、ポーズ節の直後、つまりポーズ開始から 500ms 以内にあいづちが挿入されたコンテクストのみを正例として採用する。理由は次の通りである:

1. 先行研究 [5, 4] ではポーズ節末にあいづち挿入に関する韻律特徴があるとされている。
2. 統計学習に必要となる負例の収集のためには音声データに何らかの区切りを導入する必要がある。

逆に、ポーズ節末より500ms以外にあいづちが挿入されたコンテクストは正例とはしない。

正例以外を全て負例とするか　あいづちは任意の応答とされているため、あいづちコンテクストには正例・負例コンテクスト以外の、あいづち挿入確率があまり高くないグレイゾーンが存在すると考えている。よって、正例を除いたコンテクストを全て負例に分類することは好ましくない。

もし、個人の癖や会話への参加態度に依存しない、一般的なあいづちコンテクストがあるとすれば、それはより多くの人間が共通して反応するようなコンテクストだと考えるのが妥当である。そこでこのようなコンテクストを特定するために、本研究では収録した音声対話コーパスから抜粋した音声刺激を複数の被験者に提示してそのあいづち応答を観測し、あいづちの正例・負例コンテクストを特定する実験を試みる。

3.2　実験方法

実験はおおまかに次のようなものである:

1. 音声対話コーパス [7] 中の発話音声を発話単位に切り分け、実験刺激を作る。
2. 被験者に音声刺激を提示し、キー入力であいづちを打ってもらう。被験者にはなるべく多めにあいづちを打つように頼んだ。
3. 音声刺激とあいづち挿入箇所の対応を取る。
4. あいづちが打たれた地点より500ms先行する音声部分を仮のあいづちコンテクストとする。
5. おのおののあいづちコンテクストに対してあいづちを打った被験者の人数のヒストグラムを作る。

被験者は22 ～ 30歳の奈良先端大の学生18人である。実験刺激は176 set準備し、その中からランダムに88set選んで被験者1人あたりの刺激とした。1つの実験刺激に対して9人の被験者が割り当たるようにした。また、刺激の提示順序はランダムに行ない、元の対話の文脈情報は利用できないようにした。実験刺激はそれぞれ複数のポーズ節で構成されており、あいづちを挿入された直前のポーズ節があいづちコンテクストとなる。

3.3　分析方法

実験刺激176setに対して被験者の打ったあいづちの延べ人数をあいづちコンテクストごとにカウントしてヒストグラムを取り、より多くの人があいづちを打ったコンテクストを正例、逆にほとんど打たれなかったコンテクストを負例とする。

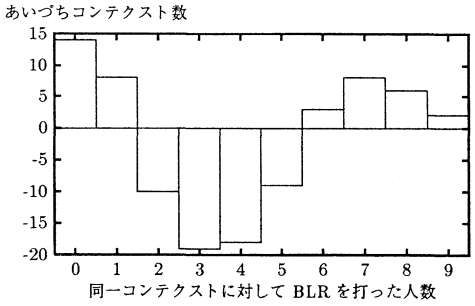


図 1: BLR の分布 (二項分布との差)

もちろん、これらのあいづちは実際のあいづちとは収録された環境、会話への集中度が違い、また文脈の欠如等もあるため、すぐに同等のものと看做すことはできない。よって本論文では心理実験により得られるこれらのあいづちをBLR (Backchannel Like Response) と呼び、コーパス中に現れたあいづちと区別する。

3.4　実験結果

BLRのヒストグラムを図1に示す。この図はBLRの分布の偏りを分かりやすくするため、あいづちの有無に対する二項分布 (平均値3.1, 総コンテクスト数=204, 総BLR数=1197) との差分で示す。

以上の結果からBLRの分布が二項分布からかけ離れていることがわかり、次のことが言える。

1. 9人中8人以上が共通してあいづちを打たないコンテクストが多数存在する。
2. 9人中6人以上が共通してあいづちを打つコンテクストが多数存在する。
3. 2人から5人があいづちを打つコンテクストは少ない。

この結果を基にコンテクストを以下の3群に分けることにする。

正例群 (6-9人) 107例
中間群 (2-5人) 132例
負例群 (0-1人) 98例

3.5　コーパス中のあいづちとの比較

BLRと同様に、音声対話コーパス中のあいづちに対してもこれらを500msを超えない範囲で先行するポーズ節をあいづちコンテクストとして数え、それぞれのコンテクストがBLRの3群のどれに対応するかを調べたところ、次のような結果が得られた。

正例群 (6-9 人) 33 例
中間群 (2-5 人) 8 例
負例群 (0-1 人) 2 例

この結果から実際のあいづちも BLR の正例群に対応するものが多く、逆に負例に対応するものは少ないことがわかる。このことは、実験的に得られた BLR が実際の会話中のあいづちと概ね似たコンテクストで挿入されることを示している。そこで、以降の統計的学習では心理実験によって得られた BLR の正例群と負例群を用いてあいづちコンテクストの学習を行なう。

4 あいづち挿入箇所の推定

本節では 3 節の分析により同定されたあいづちの正例・負例のコンテクストを判別する韻律特徴を機械的に学習し、これに基づいて収録した音声対話コーパスを用いたあいづち挿入のシミュレーション実験を行ない、判別能力を検証する。

4.1 着目する韻律特徴

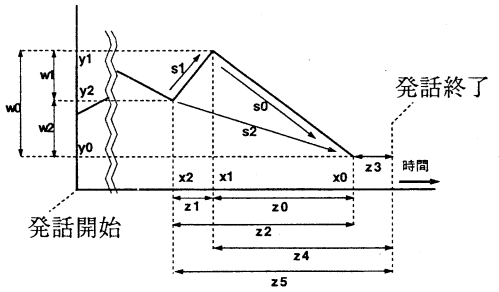
先行研究では、あいづち推定の手がかりとして 2 節に挙げたような韻律的・言語的特徴に注目している。本研究では機械的に抽出するのが容易な韻律情報だけを用いてあいづちの推定を行なうことを目的にしている。よって、語彙的な特徴 (例: 終助詞や間投助詞の「ね」) や統語的な特徴 (例: 文末)、非言語行動 (例: 頭の動き) などは用いない。また、あいづちコンテクストとして、ポーズ直前 (ポーズ節末尾) の特徴に着目する。

これらを踏まえ、本研究では以下の特徴を用いる。

- ポーズ節全体の継続長
- ポーズ節末における基本周波数, パワー曲線の極大・極小の位置と値
- ポーズ節末における基本周波数, パワー曲線に見られる上昇、下降の勾配の強さとその幅

4.2 学習データの作成

前節で述べた韻律特徴を抽出する。まず BLR の分析によって得られた正例群と負例群の各ポーズ節から本研究であいづちコンテクストと考えているポーズ前の 500ms の区間を切り出し、音響分析ソフトウェア ESPS[1] を用いて基本周波数、パワーを抽出する。抽出した基本周波数・パワーの包絡線を最小二乗法を用いて 3 次関数で近似し、さらに 2 つの変曲点と 500ms 区間の両端を結んだ 3 本の折れ線で近似して使用する。用いた特徴の一覧を図 2 に示す。



x0 ポーズ節末からスキャンして有声確率が 0.5 を上回った位置 (学習データとしては使用せず)
x1 包絡線の変曲点 1 の位置 (学習データとしては使用せず)
x2 包絡線の変曲点 2 の位置 (学習データとしては使用せず)
y0, y1, y2 x0, x1, x2 におけるピッチ・パワーの値
z0, z1, z2 x0 と x1, x1 と x2, x0 と x2 の間の時間
z3, z4, z5 SU 末から x0, x1, x2 までの時間 (ピッチのみ)
z6, z7, z8 z0 と z2, z1 と z2 の時間の比, z0 と z1 の時間の差 (z6=z0/z2, z7=z1/z2, z8=z0/z1)
w0, w1, w2 y0 と y1, y1 と y2, y0 と y2 の差分
s0, s1, s2 x0 と x1, x1 と x2, x0 と x2 の包絡線の勾配
t0, t1 包絡線の変曲の程度 (t0=s1/s0, t1=s1−s0)
full_length ポーズ節全体の長さ

図 2: 注目する韻律特徴

4.3 学習

決定木学習によりあいづちコンテクストにおける韻律的な手がかりを求める。決定木学習ソフトウェア C4.5[9] を用いて学習を行ない、表 1 のルールに変換した決定木が得られた。決定木はルール 2, 14, 15, 13, 1 の順に適用され、どのルールも充足しない場合はデフォルトルールが適用される。この決定木の inside でのあいづち推定の精度は表 2 のようになった。

この結果は予め人手で区切ったポーズ節に限定した評価である。よって関連研究である Ward[11] および岡登ら [8] との正しい比較は出来ないが、Ward の結果 (再現率 53%、適合率 33%) および岡登らの結果 (再現率 77%、適合率 49%) に比べても遜色がない。

4.4 評価実験

次に 176 個のあいづちコンテクストを 9 割の学習データと 1 割のテストデータに分割して、outside データに対する評価を行なう交差検証 (cross validation) を行なったところ、平均精度は 73.0%であった。

4.5 考察

判別ルール (表 1) から次のことがわかる:

1. 正例コンテクストを推定するルール (15, 13) により、上昇・下降イントネーションおよび上昇イン

表 1: 学習により得られた分類ルール

#	韻律特徴	判定	正解率
2	$z2_{pitch} \leq 376$ms $w2_{pitch} > -73.71$Hz full_length ≤ 724ms	挿入不可能	98.1%
14	$y0_{pitch} > 105.2$Hz $s0_{power} \leq -40.41$	挿入不可能	91.7%
15	$y0_{pitch} > 105.2$Hz $t0_{pitch} \leq -0.5712$ $s0_{power} > -40.41$ full_length ≤ 724ms	挿入可能	100%
13	$y0_{pitch} \leq 105.2$Hz $t0_{pitch} > -0.4337$ full_length > 724ms	挿入可能	90.6%
1	$w2_{pitch} > -73.71$Hz	挿入可能	70.0%
他		挿入可能	55.4%

表 2: あいづち推定の精度 (inside)

	再現率	適合率
正例	96.94%	74.22%
負例	68.87%	96.05%

正例・負例を合わせた正解率 = 82.35%

トネーションがあるとあいづちが打たれやすい

2. 負例コンテクストを推定するルール (2, 14) により平坦で短い発話の場合にあいづちが打たれにくい
3. デフォルトルールは「あいづちを打つ」
4. デフォルトルールでの適合率が著しく低い (60.9%)

1, 2 の結果は概ね先行研究によるものを支持するが、負例コンテクストについては従来それほど明確には述べられていなかったものの、本研究でその重要性が確認された。また、3 よりあいづちは打つことがデフォルトであることもわかった。これらを総合すると以下のようになる。

- あいづちを打ってはいけない場所 (負例コンテクスト) が存在する。
- 聞き手があいづちを打つことが期待されているようなコンテクスト (正例コンテクスト) が存在する。
- それ以外の部分ではあいづちを打っても打たなくてもよい。

一方で、4 の結果から負例コンテクストを判定するルールがこの他に存在すると考えられる。学習に用いる韻律特徴について今後さらに検討を加えたい。

5 おわりに

本論文は韻律情報を用いたあいづち挿入可能箇所推定法を提案し、シミュレーション実験ながら比較的高い精度で推定を実現した。今後はさらに多くのデータを用いて本手法の限界を検証する。また、本研究ではポーズ節の区切は手作業で行なっていたが、今後はポーズ挿入箇所の予測についても研究を進め、これらの成果を基に近い将来実時間あいづち応答システムとして実装したい。

謝辞

本研究に対し日頃からご指導いただいている奈良先端大の松本教授に感謝します。

参考文献

[1] Entropic Research Laboratory Inc. *Introduction to the Entropic Signal Processing System (ESPS)*, 1991.

[2] 堀口純子. 日本語教育と会話分析. くろしお出版, 1997.

[3] 今石幸子. 話し手の発話とあいづちの関係について. 大阪大学日本語学報, 13, pp. 107–120, 1994.

[4] 小磯花絵, 堀内靖雄, 土屋俊, 市川熹. 下位発話単位の音声的特徴と「あいづち」との関連について. 人工知能学会研究会資料, SIG-J-9501, pp. 9–16, 1995.

[5] メイナード泉子. 会話分析. くろしお出版, 1993.

[6] 向後千春, 山西潤一. あいづち留守番電話の試作. 日本認知科学会第 8 回大会発表論文集, pp. 72–73, 1991.

[7] 野口広彰. 韻律情報に基いたあいづち挿入箇所の推定. 修士論文, 奈良先端科学技術大学院大学, 1998.

[8] 岡登洋平, 加藤佳司, 山本幹雄, 板橋秀一. 韻律パターンを用いた相槌挿入とその評価. 情報処理学会研究報告, 96-SLP-10, pp. 33–38, 1996.

[9] Ross J. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, 1992.

[10] 杉藤美代子. 効果的な談話とあいづちの特徴及びそのタイミング. 日本語学, 12(4), pp. 11–20, 1993.

[11] Nigel Ward. In Japanese a low pitch means "Back-channel feedback please". 情報処理学会研究報告, 96-SLP-11, pp. 7–12, 1996.